**Big Data Analytics for Competitive Advantage**

ITCS/DSBA/ITIS-6100

Managerial Project Report

By

**Group 4**

*Bhavya Devineni*

*Kalyan Purushotham Karra*

*Nevetha Balasubramanian*

*Praveen Kumar Govindaraj*

*Raghunath Viswanath Nallapeta*

*Sridevi Rachakulla*

*Sushma Venkatesh Reddy*

*Vishwanarayan Sriganesh*

**SUMMARY:**

We gathered data pertaining to Revenue for the Fortune 500 companies and Election sponsorship since 1952 till 2008 and then the complete data was checked for errors and consistency post consolidation and formatting. We performed Regression Analysis, Clustering, Visualization and Topic modeling to analyze the data gathered and to establish relationship amongst the variables considered using IBM’s SPSS, Tableau and Microsoft Excel.

Also, we have established a strong relationship for all the fortune 500 companies and their contribution to the political parties of REP and DEM for the corresponding presidential election years to determine the actual values. Thus, performing statistical analysis and observation on the all the variables and attributes involved, it helped us to find out the effect of the Presidential Election results on the Fortune 500 companies Revenue and Patent count produced during those corresponding years.

The CSR data collected in first phase was taken as input for clustering to find which company is performing well with CSR versus their revenue. Cluster Analysis using R-studio facilitated us in understanding that some companies with less revenue had good CSR score and some with low CSR score had good revenue! This was useful in finding which company focuses on profits and which on social responsibilities. Also, we were able to separately cluster out (market segmentation) companies that were most and least contributing to CSR year wise based on the clustering analysis. Topic Modeling was performed on K10 reports of the organizations to know the area of focus. After Topic Modeling on IBM and Walmart K10 reports, we concluded that Walmart focuses more on retail, distribution, customers and products. On the other hand, IBM focused more on technology, data, software and cloud.

Using Tableau, we further progressed on visual analysis of the patent and revenue data of top 5 organizations. We visualized to see which state had more revenue before and after elections. We mapped these top 5 organizations visually to see how they perform with respect to Patent Count and Revenue for the Presidential Election Years. Also, we had to improvise our visualization for better presentation.

**ASSUMPTIONS:**

* Patent count for the Fortune 500 companies hold good only for the years 2008 - 2014.
* For better visualization of CSR for the Cluster analysis, we have considered only three attributes (CRS Index, Charity and Community).

**CHALLENGES:**

* The K-10 reports that had been collected don't provide much info related to Analytics and Innovation.
* Matching the latest gathered Revenue and Election data’s format to the already available data was challenging and time taking

**CONCLUSION:**

* Patent Data has a positive correlation with Revenue data, but no relation withElection results
* Election results are not an influential factor in the performance of the organizations.